

The Composite Marginal Likelihood (CML) Estimation of Panel Ordered-Response Models

Rajesh Paleti

Parsons Brinckerhoff

One Penn Plaza, Suite 200

New York, NY 10119

Phone: 512-751-5341

Email: paletir@pbworld.com

and

Chandra R. Bhat*

The University of Texas at Austin

Dept of Civil, Architectural and Environmental Engineering

301 E. Dean Keeton St. Stop C1761, Austin TX 78712

Phone: 512-471-4535, Fax: 512-475-8744

Email: bhat@mail.utexas.edu

*corresponding author

Original version: July 27, 2010

1st Revision: December 22, 2011

2nd Revision: November 30, 2012

ABSTRACT

In the context of panel ordered-response structures, the current paper compares the performance of the maximum-simulated likelihood (MSL) inference approach and the composite marginal likelihood (CML) inference approach. The panel structures considered include the pure random coefficients (RC) model with no autoregressive error component, as well as the more general case of random coefficients combined with an autoregressive error component. The ability of the MSL and CML approaches to recover the true parameters is examined using simulated datasets. The results indicate that the performances of the MSL approach (with 150 scrambled and randomized Halton draws) and the simulation-free CML approach are of about the same order in all panel structures in terms of the absolute percentage bias (APB) of the parameters and econometric efficiency. However, the simulation-free CML approach exhibits no convergence problems of the type that affect the MSL approach. At the same time, the CML approach is about 5-12 times faster than the MSL approach for the simple random coefficients panel structure, and about 100 times faster than the MSL approach when an autoregressive error component is added. As the number of random coefficients increases, or if higher order autoregressive error structures are considered, one can expect even higher computational efficiency factors for the CML over the MSL approach. These results are promising for the use of the CML method for the quick, accurate, and practical estimation of panel ordered-response models with flexible and rich stochastic specifications.

Keywords: Ordered-response model, simulated likelihood, composite marginal likelihood, cross-sectional model, panel model

1. INTRODUCTION

Ordinal discrete data arise in several empirical contexts, including ratings data (of consumer products, bonds, credit evaluation, movies, *etc.*), or likert-scale type attitudinal/opinion data (of air pollution levels, traffic congestion levels, school academic curriculum satisfaction levels, teacher evaluations, *etc.*), or grouped data (such as bracketed income data in surveys or discretized rainfall data).

Several of these applications have modeled the case of either repeated ordinal choice data (such as would be obtained from a stated preference exercise in which each respondent is asked to provide, at the same cross-sectional point in time, her/his opinion of a product multiple times based on varying the attributes of the product) or panel-based ordinal data (similar to repeated choice data, except that these are actual revealed choices made by individuals over a period of time). In this paper, the focus is on the latter case because restricted versions of the models for panel data may be applied to repeated choice data. Within this panel context, the norm in the literature is to introduce random effects and/or random parameter heterogeneity to accommodate panel effects. Such terms lead to integration in the likelihood function during estimation, resulting, in general, in the need to use numerical simulation techniques based on a maximum simulated likelihood (MSL) approach (for example, see Bhat and Zhao, 2002, Greene, 2005, Greene and Hensher, 2010) or a Bayesian inference approach (for example, see Müller and Czado, 2005, Girard and Parent, 2001). However, such simulation-based approaches can become infeasible for some panel model specifications and for long panel data. Even if feasible, the numerical simulation methods can be time-consuming and can lead to convergence problems during estimation. For instance, Bhat *et al.* (2010a) find that standard classical MSL approaches can be imprecise and have poor convergence properties, and Müller and Czado (2005) find that standard Bayesian MCMC approaches can be problematic for panel ordered response model estimations due to bad convergence properties. As a consequence, another inference approach that has seen some use recently is the simulation-free composite marginal likelihood (CML) approach. This is an estimation technique that is gaining substantial attention in the statistics field, though there has relatively little coverage of this method in econometrics and other fields. The CML method, which belongs to the more general class of composite likelihood function approaches, is based on forming a surrogate likelihood function that compounds much easier-to-compute, lower-dimensional, marginal likelihoods. Under usual regularity assumptions, and

based on the theory of estimating equations (see Lindsay, 1988, Cox and Reid, 2004), the CML estimator is consistent and asymptotically normal distributed (this is because of the unbiasedness of the CML score function, which is a linear combination of proper score functions associated with the marginal event probabilities forming the composite likelihood). The maximum CML estimator should lose some efficiency from a theoretical perspective relative to a full likelihood estimator (if this is feasible), but this efficiency loss appears to be empirically small (see Zhao and Joe, 2005, Lele, 2006, and Joe and Lee, 2009).¹ Besides, the MSL approach also loses efficiency since it involves simulation of the true analytically intractable likelihood function (see McFadden and Train, 2000). Moreover, there is always some simulation bias in the MSL method for finite number of simulation draws, and the consistency of the MSL method is guaranteed only when the number of simulation draws rises faster than the square root of the sample size (Lee, 1995 and McFadden and Train, 2000). Overall, the CML approach has some appealing properties relative to simulation techniques: It is consistent, represents a conceptually, pedagogically, and implementationally simpler procedure, and has the advantage of reproducibility of results.

The focus of this paper is on comparing the performance of the maximum-simulated likelihood (MSL) approach with the composite marginal likelihood (CML) approach in panel ordered-response situations when the MSL approach is feasible.² We use simulated data sets with known underlying model parameters to evaluate the two estimation approaches. The ability of the two approaches to recover model parameters is examined, as is the sampling variance and the simulation variance of parameters in the MSL approach relative to the sampling variance in the CML approach. The computational costs of the two approaches are also presented.³

¹ A handful of studies (see Hjort and Varin, 2008; Mardia *et al.*, 2009; Cox and Reid, 2004) have also theoretically examined the limiting normality properties of the CML approach, and compared the asymptotic variance matrices from this approach with the maximum likelihood approach. However, such a precise theoretical analysis is possible only for very simple models, and becomes much harder for models such as a panel ordered-response system.

² Note that our discussions in the paper for the panel ordered-response situation are immediately applicable to the panel binary response situation, because the latter is but a special case of the former.

³ In this paper, we do not consider the method of simulated moments (MSM) for panel ordered-response models. But, just like the MSL method, MSM also becomes cumbersome and can be fraught with convergence issues when the number of dimensions of integration involved in the likelihood function increases. Besides, in general, the MSM estimator is not as efficient as the MSL estimator. However, for some specifications such as the random effects autoregressive structure in Section 2.3, the MSM estimator has been found to have small sample properties better than the MSL estimator for a panel binary probit model and a panel multinomial probit model (especially when the autoregressive parameter takes a high value; see Keane, 1994 and Geweke *et al.*, 1997). We leave an examination of

The rest of this paper is structured as follows. In the next section, we present alternative model structures for panel ordered-response models, and discuss the maximum simulated likelihood (MSL) estimation method and the maximum CML estimation methods in the context of each of the alternative panel structures. Section 3 presents the experimental design for the simulation experiments. Section 4 presents the performance measures used for the comparison of the MSL and CML approaches, while Section 5 discusses the results. Section 6 concludes the paper by highlighting the important findings.

2. MODEL STRUCTURE

Let q be an index for individuals ($q = 1, 2, \dots, Q$), and let j be an index for the j^{th} observation (say at time t_{qi}) on individual q ($j = 1, 2, \dots, J$, where J denotes the total number of observations on individual q).⁴ Let the observed discrete (ordinal) level for individual q at the j^{th} observation be m_{qj} (m_{qj} may take one of K values; *i.e.*, $m_{qj} \in \{1, 2, \dots, K\}$). In the usual random-effects ordered response framework notation, we write the latent variable (y_{qj}^*) as a function of relevant covariates as:

$$y_{qj}^* = \boldsymbol{\beta}_q' \mathbf{x}_{qj} + \varepsilon_{qj}, y_{qj} = m_{qj} \text{ if } \psi^{m_{qj}-1} < y_{qj}^* < \psi^{m_{qj}}, \quad (1)$$

where \mathbf{x}_{qj} is a $(H \times 1)$ -vector of exogenous variables (including a constant), $\boldsymbol{\beta}_q$ is an individual-specific $(H \times 1)$ -vector of coefficients to be estimated that is a function of unobserved individual attributes, ε_{qj} is a standard normal (or logistic) error term uncorrelated across individuals q (but it may be correlated across observations j ($j = 1, 2, \dots, J$) of the same individual, depending upon the analyst's specification) , and $\psi^{m_{qj}}$ is the upper bound threshold for discrete level m_{qj}

the performance of the MSM and the CML estimators, in situations where the MSM estimator is feasible to implement for panel ordered-response models, to a future research effort.

⁴ We assume here that the number of panel observations is the same across individuals. Extension to the case of different numbers of panel observations across individuals does not pose any substantial challenges, and will be discussed later.

($\psi^0 < \psi^1 < \psi^2 \dots < \psi^{K-1} < \psi^K$; $\psi^0 = -\infty, \psi^1 = 0, \psi^K = +\infty$).⁵ Assume that the β_q vector in Equation (1) is a realization from a multivariate normal distribution $\phi(\beta)$ with a mean vector \mathbf{b} and covariance matrix $\Omega = \mathbf{L}\mathbf{L}'$, where \mathbf{L} is the lower-triangular Cholesky factor of Ω . Also, assume that the ε_{qj} term, which captures the idiosyncratic effect of all omitted variables for individual q at the j^{th} choice occasion, is independent of the elements of the β_q and \mathbf{x}_{qj} vectors. We now discuss four different model structures, based on different assumptions about the β_q vector.

2.1 Random-Effects Model

The simplest panel model is one that includes an individual-specific constant term, but does not consider heterogeneity in other parameters in β_q across individuals q . Thus, we write $\beta_q' \mathbf{x}_{qj} = \alpha_q + \gamma' \mathbf{z}_q$, where the vector \mathbf{z}_q now includes all the variables but no constant, and γ is a fixed coefficient vector to be estimated. Substituting this expression in Equation (1), and writing α_q in random effects form as $\alpha_q = \alpha + \eta_q$, we get the following equation:

$$y_{qj}^* = \alpha + \gamma' \mathbf{z}_{qj} + \eta_q + \varepsilon_{qj}, y_{qj} = m_{qj} \text{ if } \psi^{m_{qj}-1} < y_{qj}^* < \psi^{m_{qj}} \quad (2)$$

η_q in the above equation is an individual-specific random term that generates a correlation in the propensity across all of individuals q 's J observed choice occasions. It is typical to consider the heterogeneity term η_q to be normally distributed. However, other distributions may also be empirically tested, such as the logistic distribution with fatter tails. But the consideration of a normally distributed η_q with a standard normally distributed ε_{qj} is natural and convenient here, which is what we will assume. The result is the standard textbook random-effects ordered-

⁵ The model can be generalized in many ways, though the model as written is the most familiar and common panel version of the ordered-response model. For instance, the mean vector \mathbf{b} of β_q can be a function of observed individual attributes. However, this can be accommodated without any complications by redefining \mathbf{x}_{qj} to include interaction terms. Also, one can label the coefficient vector on the \mathbf{x}_{qj} variable vector as β_{qj} (rather than β_q) to allow for fixed or random elements of response specific to period j , and to capture random heterogeneity in response across individuals and choice occasions (see Bhat and Sardesai, 2006). This relabeling also then allows observed individual and choice occasion specific variable effects to be introduced in the covariance matrix of β_{qj} . Finally, one can also include heterogeneity in the variance of ε_{qj} and accommodate heterogeneity in the thresholds $\psi^2, \psi^3, \dots < \psi^{K-1}$ through careful parameterizations to ensure the ordinality conditions on the thresholds (see Eluru *et al.*, 2008). However, all these generalizations cause an explosion in parameters, and need very rich data sets to estimate parameters.

response model, which takes the same form as the random-effects binary choice model proposed by Butler and Moffitt (1982).

2.1.1 Maximum Simulated Likelihood (MSL) Estimation of Random-Effects Model

The MSL estimation of the random-effects model is relatively straightforward. The probability of the observed vector m_q of the sequence of ordinal choices $(m_{q1}, m_{q2}, m_{q3}, \dots, m_{qJ})$ for individual q , conditional on the heterogeneity term η_q , can be written as:

$$\text{Prob}(m_q) | \eta_q = \prod_{j=1}^J \left\{ \Phi(\psi^{m_{qj}} - \alpha - \gamma' z_{qj} - \eta_q) - \Phi(\psi^{m_{qj}-1} - \alpha - \gamma' z_{qj} - \eta_q) \right\} \quad (3)$$

The unconditional likelihood of the observed choice sequence is obtained by integrating out the term η_q :

$$L_q(\psi, \alpha, \gamma, \sigma) = \int_{v=-\infty}^{\infty} \left[\prod_{j=1}^J \left\{ \Phi(\psi^{m_{qj}} - \alpha - \gamma' z_{qj} - \sigma v) - \Phi(\psi^{m_{qj}-1} - \alpha - \gamma' z_{qj} - \sigma v) \right\} \phi(v) dv \right] \quad (4)$$

where $v = \frac{\eta_q}{\sigma}$, $\eta_q \sim N(0, \sigma^2)$, ψ is the vector of all threshold bounds, $\Phi(\cdot)$ is the univariate standard normal cumulative distribution, and $\phi(\cdot)$ is the corresponding univariate standard normal density function. Finally, the log-likelihood function may be written as:

$$\log L(\psi, \alpha, \gamma, \sigma) = \sum_q \log L_q(\psi, \alpha, \gamma, \sigma) \quad (5)$$

The log-likelihood function above can be maximized using Gauss-Hermite Quadrature or using a simulation method. Since, the function entails only a one dimensional integral, estimation is generally very fast and there is no convergence-related problems.

2.1.2 Composite Marginal Likelihood (CML) Estimation of Random-Effects Model

The composite marginal likelihood (CML) estimation approach (see Varin, 2008, Varin *et al.*, 2011, and Bhat *et al.*, 2010a for good reviews) is a relatively simple approach that can be used when the full likelihood function is cumbersome or plain infeasible to evaluate due to the underlying complex dependencies, as is the case with certain specifications of panel models that entail high dimensional integration in the likelihood function. While there have been recent

advances in simulation techniques within a classical or Bayesian framework that assist with such model estimation situations (see Bhat, 2003, Beron and Vijverberg, 2004, and LeSage, 2000), these techniques are impractical and/or infeasible in situations in some panel ordered-response situations (see, for example, Varin and Czado, 2010). Further, even when the integration is of low dimension, the CML method may have a substantial edge in terms of computation speed. The CML method, which belongs to the more general class of composite likelihood function approaches (see Lindsay, 1988), is based on forming a surrogate likelihood function that compounds easier-to-compute, lower-dimensional, likelihoods of marginal events. In panel data, the simplest CML, formed by assuming independence across observed choice instances from the same individual, entails the product of univariate densities (for continuous data) or probability mass functions (for discrete data). However, this approach does not provide estimates of dependence among the individual observations. Another approach is the pairwise likelihood function formed by the product of power-weighted likelihood contributions of all or a selected subset of couplets (*i.e.*, pairs of observed events). This pairwise method corresponds to a composite marginal approach based on bivariate marginals. For individual q , the pairwise likelihood function is:

$$L_{CML,q}(\boldsymbol{\psi}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\sigma}) = \left(\prod_{j=1}^{J-1} \prod_{g=j+1}^J [\Pr(y_{qj} = m_{qj}, y_{qg} = m_{qg})] \right)^{w_q}, \quad (6)$$

where w_q is a power weight to be chosen based on efficiency considerations (see Kuk and Nott, 2000, Zhao and Joe, 2005, Joe and Lee, 2009). When the number of choice occasions are the same across individuals, as we assume in the current paper, this power weight term may be ignored and arbitrarily set to one for each individual. When the number of choice occasions are different across individuals, setting w_q to be one for all individuals will give more weight to individuals who have more choice occasions than to individuals who have fewer choice occasions. In this situation, the weights to be used are discussed in Section 2.5.⁶

⁶ A note about the CML method is in order here. The CML method is helpful when there are multiple observed events, with the events being dependent because of covariance in the underlying latent variables generating the events. This is the case when one is dealing with panel choice data, or with the case of cross-sectional choice data with spatial and/or social dependence across the choice instances of decision agents, or with combinations of the two. In these situations, the multivariate probability of the dependent events in the likelihood function is replaced in the CML approach by a compounding of the probabilities of lower dimensional events (such as the likelihood of each pair of observed choice instances from the same decision agent in panel data, or the likelihood of each pair of observed choice instances of different decision agents in dependent cross-sectional data). Of course, the CML

To write the pairwise likelihood function in terms of the parameters to be estimated in the simple random-effects model, note that the joint distribution of the latent variables $(y_{q1}^*, y_{q2}^*, \dots, y_{qJ}^*)$ for the q^{th} individual is multivariate normal with standardized mean vector

$$\frac{\alpha + \gamma' z_{q1}}{\mu}, \frac{\alpha + \gamma' z_{q2}}{\mu}, \dots, \frac{\alpha + \gamma' z_{qJ}}{\mu} \text{ and a correlation matrix with constant non-diagonal entries } \frac{\sigma^2}{\mu^2},$$

where $\mu = \sqrt{1 + \sigma^2}$. Then, we can write

$$\begin{aligned} \Pr(y_{qj} = m_{qj}, y_{qg} = m_{qg}) &= \Phi_2(\delta^{m_{qj}}, \delta^{m_{qg}}, \rho_{jg}) - \Phi_2(\delta^{m_{qj}}, \delta^{m_{qg}^{-1}}, \rho_{jg}) - \Phi_2(\delta^{m_{qj}^{-1}}, \delta^{m_{qg}}, \rho_{jg}) \\ &\quad + \Phi_2(\delta^{m_{qj}^{-1}}, \delta^{m_{qg}^{-1}}, \rho_{jg}), \end{aligned} \quad (7)$$

$$\text{where } \delta^{m_{qj}} = \frac{\psi^{m_{qj}} - \alpha - \gamma' z_{qj}}{\mu}, \mu = \sqrt{1 + \sigma^2}, \rho_{jg} = \frac{\sigma^2}{\mu^2}$$

The logarithm of the pairwise likelihood function is:

$$\log L_{\text{CML}}(\psi, \alpha, \gamma, \sigma) = \sum_q \sum_{j=1}^{J-1} \sum_{g=j+1}^J \log(\Pr(y_{qj} = m_{qj}, y_{qg} = m_{qg})) \quad (8)$$

The CML estimator $\hat{\theta} = (\psi, \alpha, \gamma', \sigma)'$ obtained by maximizing the above function is consistent and asymptotically normally distributed with the asymptotic variance matrix vector given by the inverse of the Godambe's (1960) sandwich information matrix.⁷

2.2 Random Coefficients Model

In this model, the coefficients on the exogenous variables are also considered to be randomly distributed. Going back to Equation (1), assume that β_q is multivariate normal distributed with

cannot break down a multivariate probability expression in cases where individual events themselves are of high dimension. For instance, consider a cross-sectional multinomial probit model with independent probabilities across decision agents. When the number of alternatives involved is high, the probability of choice for each individual takes the form of a high multivariate normal cumulative distribution (MVNCD) function. In this case, the MVNCD function evaluation for each agent has to be evaluated using simulation or other analytic approximation techniques (see Bhat, 2011 for a detailed discussion).

⁷ The random effects specification is considered here just by way of introducing the CML method for a simple specification. Given the single dimensional integral in the full likelihood function for the random effects ordered-response model, there is really no reason to use the CML estimator in this random effects case. Indeed, the CML estimator involves developing all pairwise probabilities, which, in general, will be more time consuming than using fast quadrature or other techniques on the full likelihood function. However, the situation becomes different for the other cases discussed next.

mean vector \mathbf{b} and covariance $\mathbf{\Omega} = \mathbf{L}\mathbf{L}'$.⁸ For later use, define $\boldsymbol{\beta}_q = \mathbf{b} + \tilde{\boldsymbol{\beta}}_q$, where $\tilde{\boldsymbol{\beta}}_q$ is multivariate normal distributed with a mean vector of zeros and a covariance matrix given by $\mathbf{\Omega}$. Note that it is not necessary that all elements of the $\boldsymbol{\beta}_q$ be random. That is, the analyst may specify fixed coefficients on some exogenous variables in the model, though it will be convenient in presentation to assume that all elements of $\boldsymbol{\beta}_q$ are random.

2.2.1 Maximum Simulated Likelihood Estimation

The likelihood function contribution of individual q for the random coefficients model is:

$$L_q(\boldsymbol{\psi}, \mathbf{b}, \mathbf{\Omega}) = \int_{\boldsymbol{\beta}=-\infty}^{\infty} \left[\prod_{j=1}^J \left\{ \Phi(\boldsymbol{\psi}^{m_{qj}} - \boldsymbol{\beta}' \mathbf{x}_{qj}) - \Phi(\boldsymbol{\psi}^{m_{qj}-1} - \boldsymbol{\beta}' \mathbf{x}_{qj}) \right\} f(\boldsymbol{\beta} | \mathbf{b}, \mathbf{\Omega}) d\boldsymbol{\beta} \right] \quad (9)$$

where $f(\cdot)$ is multivariate normal density function with mean vector \mathbf{b} and covariance $\mathbf{\Omega}$

The log-likelihood function is:

$$\log L(\boldsymbol{\psi}, \mathbf{b}, \mathbf{\Omega}) = \sum_q \log L_q(\boldsymbol{\psi}, \mathbf{b}, \mathbf{\Omega}) \quad (10)$$

⁸ One may also consider non-normal distributions for the random coefficients. In general, doing so does not allow the analytic development of the joint distribution of the latent variables $(y_{q1}^*, y_{q2}^*, \dots, y_{qJ}^*)$, which is needed for the use of the composite marginal likelihood (CML) method (normality, on the other hand, leads to a multivariate normal distribution for the latent variables because of the conjugate addition property of the normal distribution). On the other hand, the MSL approach is generic and applicable for both normal as well as non-normal distributions. While there have been several applications of a non-normal distribution for random coefficients (such as on the cost coefficient), as indicated by Bhat (2011), it has been well known that using non-normal distributions can lead to convergence/computational problems in the MSL approach, and it is not uncommon to see researchers consider non-normal distributions only to eventually revert to the use of a normal distribution (see, for example, Bartels *et al.*, 2006 and Small *et al.*, 2005). Recently, Bhat and Sidharthan (2012) have come up with a new approach to introduce non-normality in unordered-response that still allows the use of the CML estimation approach, thus doing away with the computational problems of the MSL. Specifically, they propose the use of a multivariate skew normal (MVSN) distribution for the random coefficients. The MVSN distribution is tractable, parsimonious in parameters that regulate the distribution and its skewness, and includes the normal distribution as a special interior point case. It also is a very flexible unimodal density structure that allows a “seamless” and “continuous” variation from normality to non-normality, and can replicate a variety of smooth unimodal density shapes with tails to the left or right as well as with a high modal value (sharp peaking) or low modal value (flat plateau). At the same time, the joint distribution of the latent variables takes a MVSN form, which lends itself nicely to the use of the CML inference approach. In this paper, we use a multivariate normal distribution for the random coefficients, and leave the extension to the non-normal case (*a la* Bhat and Sidharthan, 2012) for a future effort.

The expression $\sum_q \log L_q(\boldsymbol{\psi}, \mathbf{b}, \boldsymbol{\Omega})$ entails integration of dimension equal to the number of elements of $\boldsymbol{\beta}_q$. Alternatively, one could combine the error terms $\tilde{\boldsymbol{\beta}}_q' \mathbf{x}_{qj} + \varepsilon_{qj}$ for each individual q and choice occasion j . By the conjugate addition property of the normal distribution, this composite error term is also normally distributed. Defining $\mathbf{y}_q^* = (y_{q1}^*, y_{q2}^*, \dots, y_{qJ}^*)'$ ($J \times 1$ matrix) and $\mathbf{x}_q = (x_{q1}, x_{q2}, \dots, x_{qJ})'$ ($J \times K$ matrix), we may write $\mathbf{y}_q^* \sim MVN_J[\mathbf{x}_q \mathbf{b}, (\mathbf{x}_q \boldsymbol{\Omega} \mathbf{x}_q' + \mathbf{I}_J)]$ with \mathbf{I}_J being the identity matrix of size J and MVN_J being the multivariate normal density function of dimension J . One can then integrate this J -dimensional multivariate normal vector between the appropriate multivariate threshold bounds to obtain the likelihood function as an alternative to Equation (9). Both these likelihood functions are identical; if $K < J$, then Equation (9) involves a lesser dimension of integration, while if $K > J$, the alternative form involves a lesser dimension of integration. When $K = J$ (as in our simulation exercise discussed later), both forms have the same dimensionality of integration. But the likelihood form in Equation (9) is easier since it entails integration over the entire real domain (from $-\infty$ to $+\infty$) rather than a rectangular bounded domain in the alternative form. So, we use Equation (9) as the likelihood function for the random coefficients model.

The estimation of the log-likelihood function in Equation (10) cannot, in general, be pursued using quadrature techniques due to the curse of dimensionality. Instead, it is typical to use quasi-Monte Carlo (QMC) techniques for simulation estimation (Bhat, 2001, 2003). To ensure the positive definiteness of the covariance matrix $\boldsymbol{\Omega}$, the likelihood function contribution of individual q of Equation (9) is rewritten in terms of the Cholesky-decomposed matrix \mathbf{L} of $\boldsymbol{\Omega}$. The maximum simulated likelihood approach then proceeds by optimizing with respect to the elements of \mathbf{L} rather than $\boldsymbol{\Omega}$. Once convergence is achieved, the implied covariance matrix $\boldsymbol{\Omega}$ may be reconstructed from the estimated matrix \mathbf{L} .

While there have been important advances in terms of the QMC based simulation of the mixed panel models for random coefficients, these QMC methods continue to be quite expensive for the usual sample sizes encountered in practice. Besides, even for low to moderate dimensions of integration (of the order of four to seven), the numerical simulators can lead to numerical instability, non-convergence, and imprecision problems as the number of dimensions increases. Bhat *et al.* (2010a) find another bothersome issue with these MSL simulation methods even for low to moderate dimensions in that even if the log-likelihood function is computed with good

precision, so that the simulation error in the estimated parameters is small, the computation of the numerical Hessian is not very reliable. But a good estimate of the Hessian is needed for the sandwich estimator of the covariance matrix in the MSL method (the alternative of using the inverse of the cross product of the first derivatives is not appropriate in the MSL because of simulation noise introduced when using a finite number of draws per individual, see McFadden and Train, 2000). The only way out of the problem is to compute the log likelihood function with a very high level of precision, which can lead to high computational times even at low dimensions.

2.2.2 CML Estimation

The pairwise marginal likelihood function for the random coefficients panel ordered-response model is much simpler than the full likelihood function in Equation (9), as also suggested by Renard *et al.* (2004) in the context of a panel binary choice model. In particular, based on the joint distribution of the latent variable vector $\mathbf{y}_q^* = (y_{q1}^*, y_{q2}^*, \dots, y_{qJ}^*)'$ for the q^{th} individual, one can write the contribution of the q^{th} individual to the pairwise-likelihood function as:

$$L_{CML,q}(\boldsymbol{\psi}, \mathbf{b}, \boldsymbol{\Omega}) = \left(\prod_{j=1}^{J-1} \prod_{g=j+1}^J [\Pr(y_{qj} = m_{qj}, y_{qg} = m_{qg})] \right)^{w_q} \quad (11)$$

$$= \left(\prod_{j=1}^{J-1} \prod_{g=j+1}^J \left[\Phi_2(\delta^{m_{qj}}, \delta^{m_{qg}}, \rho_{qjg}) - \Phi_2(\delta^{m_{qj}}, \delta^{m_{qg}-1}, \rho_{qjg}) \right. \right. \\ \left. \left. - \Phi_2(\delta^{m_{qj}-1}, \delta^{m_{qg}}, \rho_{qjg}) + \Phi_2(\delta^{m_{qj}-1}, \delta^{m_{qg}-1}, \rho_{qjg}) \right] \right)^{w_q},$$

where $\delta^{m_{qj}} = \frac{\boldsymbol{\psi}^{m_{qj}} - \mathbf{b}' \mathbf{x}_{qj}}{\sqrt{\text{Var}(y_{qj}^*)}}$ and $\rho_{qjg} = \frac{\text{Cov}(y_{qj}^*, y_{qg}^*)}{\sqrt{\text{Var}(y_{qj}^*)} \sqrt{\text{Var}(y_{qg}^*)}}$

In the above expression, the $\text{Var}(y_{qj}^*)$, $\text{Var}(y_{qg}^*)$ and ρ_{qjg} terms are obtained by picking off the appropriate 2×2 sub-matrix of the covariance matrix of $(y_{q1}^*, y_{q2}^*, \dots, y_{qJ}^*)$ given by $(\mathbf{x}_q \boldsymbol{\Omega} \mathbf{x}_q' + \mathbf{I}_J)$. The logarithm of the pairwise likelihood function is:

$$L_{CML}(\boldsymbol{\psi}, \boldsymbol{\beta}, \boldsymbol{\Omega}) = \sum_q \log L_{CML,q}(\boldsymbol{\psi}, \boldsymbol{\beta}, \boldsymbol{\Omega}), \quad (12)$$

The asymptotic variance expression is given by the sandwich estimator, as discussed earlier in Section 2.1.2.

The random coefficients model is commonly referred to as the mixed model in the literature, and the CML approach above is an alternative to the commonly used MSL approach. As in the MSL case, one can ensure the positive-definiteness of $\mathbf{\Omega}$ in the CML method by writing the logarithm of the pairwise-likelihood in terms of the Cholesky-decomposed elements of $\mathbf{\Omega}$ and maximizing with respect to these elements of the Cholesky factor. Essentially, this entails passing the Cholesky elements as parameters to the optimization routine, constructing the $\mathbf{\Omega}$ matrix internal to the optimization routine, and then picking off the appropriate sub-matrix for the pairwise likelihood components.⁹

2.3 Random Effects Autoregressive Structure

The standard random-effects ordered-response model of Equation (1) allows easy estimation, since there is only a one-dimensional integral for each individual. However, the assumption of equal correlation across the multiple observations on the same individual is questionable, especially for medium-to-long panels. An alternative would be to allow a time-stationary error component, but also allow serial correlation within each subject-specific series of observations (see Varin and Czado, 2010 and Bhat *et al.*, 2010a consider a similar model but for the unordered multinomial probit (MNP) model rather than the multivariate ordered-response system considered here). For instance, one may adopt an autoregressive structure of order one for the

⁹ In the random coefficient case when $K < J$, and $K \leq 3$, the MSL estimation is likely to be preferable to the CML. This is because integrating over up to three dimensions is quite fast and accurate using quasi-Monte Carlo simulation techniques. This is particularly so when J is also large, because the number of pairings in the CML is high. For the case when $K < J$ and $K > 3$, or $K \geq J > 3$, the CML is likely to become attractive, because of the MSL-related problems mentioned earlier for moderate dimensions of integration. For example, when $K = J = 5$, as in our simulation exercise discussed later, the CML is fast since it entails the evaluation of only 10 probability pairings for each individual (each pairing involving bivariate normal cumulative distribution function evaluations) rather than a five-dimensional integration for each individual in the MSL estimation. Note that one may be tempted to think that the CML loses this edge when J becomes large. For instance, when $J = 10$, there would be 45 probability pairings for each individual in the way things have been presented so far. But the surrogate likelihood function in the CML estimation can be formulated in many different ways rather than the full pairings approach presented here. Thus, one could consider only the pairing combinations of the first five (or five randomly selected) choice occasions for each individual, and assume independence between the remaining five choice occasions and between each of these remaining choice occasions and the choice occasions chosen for the pairings. Basically, the CML approach is flexible, and allows customization based on the problem at hand. The issue then becomes one of balancing between speed gain/convergence improvement and efficiency loss. Besides, the CML can also use triplets or quadruplets rather than the couplets considered here. Overall, how exactly to form a CML function remains a wide open area of empirical research (especially because a precise theoretical analysis of the properties of the CML estimator is not possible except for the simplest of models), though it is generally accepted that the full pairwise approach tends to be a good balance between estimation speed and estimator efficiency for many empirical contexts (see Bhat *et al.*, 2010a). In this paper, we do not delve into the issue of how to form the CML function, but use a full pairwise formulation. Future studies should extend the analysis to consider the CML performance under different formulations and different values of K and J .

error terms of the same individual, so that $\text{corr}(\varepsilon_{qj}, \varepsilon_{qg}) = \rho^{|t_{qj} - t_{qg}|}$ ($0 < \rho < 1$), where t_{qj} is the measurement time of observation y_{qj} . This is in addition to the equal correlation across observations of the same individual, due to the individual specific random term η_q in the Equation (2). The autoregressive error structure specification results in a joint multivariate distribution of the latent variables $(y_{q1}^*, y_{q2}^*, \dots, y_{qJ}^*)$ for the q^{th} individual with standardized mean vector $\frac{\alpha + \gamma' z_{q1}}{\mu}, \frac{\alpha + \gamma' z_{q2}}{\mu}, \dots, \frac{\alpha + \gamma' z_{qJ}}{\mu}$ and a correlation matrix Σ with entries such that $\text{corr}(y_{qj}^*, y_{qg}^*) = (\sigma^2 + \rho^{|t_{qj} - t_{qg}|}) / \mu^2$, where $\mu = \sqrt{1 + \sigma^2}$.

2.3.1 Maximum Simulated Likelihood Estimation

The random effects autoregressive model structure, while much more realistic than the simple random effects, also costs dearly in terms of computational time. In particular, rather than a single dimension of integration, we now have an integral of dimension J for individual q . The likelihood function for individual q is:

$$L_q(\psi, \alpha, \gamma, \sigma, \rho) = \Pr(y_{q1} = m_{q1}, y_{q2} = m_{q2}, \dots, y_{qJ} = m_{qJ})$$

$$L_q(\psi, \alpha, \gamma, \sigma, \rho) = \int_{w_1 = \delta^{m_{q1}-1}}^{\delta^{m_{q1}}} \int_{w_2 = \delta^{m_{q2}-1}}^{\delta^{m_{q2}}} \dots \int_{w_J = \delta^{m_{qJ}-1}}^{\delta^{m_{qJ}}} \phi_J(w_1, w_2, \dots, w_J | \Sigma) dw_1 dw_2 \dots dw_J \quad (13)$$

where $\delta^{m_{qj}} = \frac{(\psi^{m_{qj}} - \alpha - \gamma' z_{qj})}{\mu}$, $\mu = \sqrt{1 + \sigma^2}$, and ϕ_J is the J -variate standard multivariate normal

density function. The integral above may be evaluated using the Geweke-Hajivassiliou-Keane (GHK) simulator (see Geweke, 1991, Hajivassiliou and McFadden, 1998, and Keane, 1994) or the Genz-Bretz (GB) simulator (Genz and Bretz, 1999; 2002, and Mi *et al.*, 2009), which are among the most effective simulators for evaluating rectangular multivariate normal probabilities (*i.e.*, bounded as opposed to unbounded limits of integration). Positive definiteness of the correlation matrix is guaranteed as long as $\sigma > 0$, and $0 < \rho < 1$, which can be easily imposed through appropriate parameterizations.

The problems with the MSL approach for the random-effects autoregressive model are similar to the ones discussed earlier in the context of the random coefficients model. However,

while the dimension of random coefficients and therefore the dimensionality of the integration may be relatively low for the random coefficients model, the number of observations per individual, and therefore the dimensionality of integration, can be very high for the random-effects autoregressive model. For instance, in Varin and Czado (2010), the authors examine the headache pain intensity of patients at different points of time during the day and across several consecutive days. In this study, the full information likelihood estimation has of the order of 800 dimensions of integration for some of the individual-specific likelihood contributions, an infeasible proposition for model parameter estimation using any computer intensive simulation procedure.

2.3.2 Composite Marginal Likelihood Estimator

The pairwise marginal likelihood function for individual q in the random-effects autoregressive structure is:

$$L_{CML,q}(\boldsymbol{\psi}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \sigma, \rho) = \prod_{g=j+1}^J \prod_{j=1}^{J-1} \left[\begin{aligned} &\Phi_2(\delta^{m_{qj}}, \delta^{m_{qg}}, \rho_{jg}) - \Phi_2(\delta^{m_{qj}}, \delta^{m_{qg-1}}, \rho_{jg}) \\ &- \Phi_2(\delta^{m_{qj-1}}, \delta^{m_{qg}}, \rho_{jg}) + \Phi_2(\delta^{m_{qj-1}}, \delta^{m_{qg-1}}, \rho_{jg}) \end{aligned} \right], \quad (14)$$

where $\delta^{m_{qj}} = (\boldsymbol{\psi}^{m_{qj}} - \boldsymbol{\alpha} - \boldsymbol{\gamma}' \mathbf{z}_{qj}) / \mu$, $\mu = \sqrt{1 + \sigma^2}$, and $\rho_{jg} = (\sigma^2 + \rho^{|t_{qj} - t_{qg}|}) / \mu^2$.

The logarithm of the likelihood function is:

$$LL_{CML}(\boldsymbol{\psi}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \sigma, \rho) = \sum_q \log L_{CML,q}(\boldsymbol{\psi}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \sigma, \rho). \quad (15)$$

Compared to the MSL technique, the pairwise approach only entails bivariate normal distributions, which can be evaluated rapidly. The asymptotic covariance matrix may be obtained as the inverse of the Godambe sandwich information estimator, as in Section 2.1.2.

2.4 Random Coefficients Autoregressive Structure

This general structure combines the random coefficient structure with the autoregressive structure. The form of the model is as follows (using the same notations as earlier):

$$y_{qj}^* = \boldsymbol{\beta}_q' \mathbf{x}_{qj} + \varepsilon_{qj}, y_{qj} = m_{qj} \text{ if } \boldsymbol{\psi}^{m_{qj-1}} < y_{qj}^* < \boldsymbol{\psi}^{m_{qj}} \quad (16)$$

with $\boldsymbol{\beta}_q \sim N(\mathbf{b}, \boldsymbol{\Omega})$, $\varepsilon_{qj} \sim N(0, 1)$, and $\text{corr}(\varepsilon_{qj}, \varepsilon_{qg}) = \rho^{|t_{qj} - t_{qg}|}$, $0 < \rho < 1$

2.4.1 Maximum Simulated Likelihood Estimation

$$L_q(\boldsymbol{\psi}, \mathbf{b}, \boldsymbol{\Omega}, \rho) = \int_{\omega_1=\delta^{m_{q1}-1}}^{\delta^{m_{q1}}} \int_{\omega_2=\delta^{m_{q2}-1}}^{\delta^{m_{q2}}} \cdots \int_{\omega_J=\delta^{m_{qJ}-1}}^{\delta^{m_{qJ}}} \phi_J(\omega_1, \omega_2, \dots, \omega_J | \boldsymbol{\Sigma}_q) d\omega_1 d\omega_2 \dots d\omega_J \quad (17)$$

where $\delta^{m_{qj}} = \frac{(\boldsymbol{\psi}^{m_{qj}} - \mathbf{b}' \mathbf{x}_{qj})}{\sqrt{\text{Var}(y_{qj}^*)}}$, and $\boldsymbol{\Sigma}_q$ is a correlation matrix with entries such that

$$\Sigma_{qjg} = \frac{\text{Cov}(y_{qj}^*, y_{qg}^*)}{\sqrt{\text{Var}(y_{qj}^*) \text{Var}(y_{qg}^*)}}. \text{ The covariance matrix of the latent } (y_{q1}^*, y_{q2}^*, \dots, y_{qJ}^*) \text{ dependent}$$

variables is given by $(\mathbf{x}_q \boldsymbol{\Omega} \mathbf{x}_q' + \mathbf{R}_q)$ where \mathbf{R}_q is a $J \times J$ matrix with the entry $(\rho^{|t_{qj} - t_{qg}|})$ for the jg^{th} element of the matrix ($j = 1, 2, \dots, J; g = 1, 2, \dots, J$). This covariance matrix is positive definite as long as $\boldsymbol{\Omega}$ can be written as $\mathbf{L}\mathbf{L}'$ (where \mathbf{L} is the lower triangular Cholesky factor of $\boldsymbol{\Omega}$) and $0 < \rho < 1$. The log-likelihood function is finally computed as:

$$\sum_q \log L_{MSL,q}(\boldsymbol{\psi}, \mathbf{b}, \boldsymbol{\Omega}, \rho) \quad (18)$$

The MSL estimation of the random coefficients autoregressive model is particularly cumbersome, and we are not aware of an earlier study considering such an ordered-response model even though it naturally arises as a combination of random coefficients and a first-order autocorrelation process (however, Keane (1997) considered a similar model for the unordered multinomial probit (MNP) model rather than the multivariate ordered-response system considered here).

2.4.2 Composite Marginal Likelihood Estimation

The pairwise function for individual q in the random coefficients autoregressive structure takes the same form as for the random coefficients structure (see Equation 11), but with the important difference that the $\text{Var}(y_{qj}^*)$, $\text{Var}(y_{qg}^*)$ and ρ_{qjg} are obtained by picking off the appropriate (2×2) sub-matrix of the covariance matrix of $(y_{q1}^*, y_{q2}^*, \dots, y_{qJ}^*)$ given by $(\mathbf{x}_q \boldsymbol{\Omega} \mathbf{x}_q' + \mathbf{R}_q)$ (rather than $(\mathbf{x}_q \boldsymbol{\Omega} \mathbf{x}_q' + \mathbf{I}_J)$ in the random coefficients models). The logarithm of the pairwise likelihood function and the asymptotic variance expression of the estimator are obtained in the usual way.

2.5 Weights in CML Estimation

As discussed in Section 2.1.2, the pairwise likelihood function involves a weight term which must be chosen based on efficiency considerations. When the number of choice occasions is the same across individuals, the power weight w_q for each individual may be set to 1. However in the more general case when the number of observations from individual q is J_q , Le Cessie and Van Houwelingen (1994) suggest, based on their correlated binary model analysis, that each individual should contribute about equally to the CML function. This may be achieved by power-weighting each individual's CML contribution by a factor that is the inverse of the number of choice occasions minus one (in our context, this is $[J_q - 1]^{-1}$). The net result is that the composite likelihood contribution of individual q collapses to the likelihood contribution of the individual under the case of independence across choice occasions. In a recent paper, Joe and Lee (2009) theoretically studied the issue of efficiency in the context of a simple random-effects binary choice model. They indicate that the weights suggested by Le Cessie and Van Houwelingen (1994) can provide poor efficiency when the correlation between pairs of the underlying latent variables for the repeated binary choices over time is moderate to high. Intuitively, Joe and Lee's discussion is based on the concept that, when there is perfect dependence between each pair of inter-temporal binary choices, each pairing should contribute the same amount of information to the CML function. While Joe and Lee's theoretical analysis is confined to a simple random-effects binary model, it may be extended to the random-effects ordered panel case (and also to other ordered-response panel models discussed later). Joe and Lee (2009) proposed the optimal power weight for individual q in the unbalanced panel case as $w_q = (J_q - 1)^{-1}[1 + 0.5(J_q - 1)]^{-1}$. In the current paper, we ignore the weight term, since we are focusing our simulation experiments on the case of the same number of choice occasions from each individual.¹⁰

3. SIMULATION STUDY

In the current paper, we assess the performance of the CML technique for panel ordered response models in the context of the random coefficients structure and the random coefficients

¹⁰ The focus in the current paper is on comparing the performance of the maximum simulated likelihood approach with the CML approach, so we steer clear of issues related to optimal weights for the CML approach by considering the "equal observations across individuals" case.

autoregressive structure. This is because of three reasons. First, the random coefficients structure subsumes the random effects structure as a special case. Second, the random coefficients structure (without autoregressive error terms) has been extensively used in binary and ordered response modeling (for example, see Bhat and Zhao, 2002, Greene, 2000). The use of the CML technique can lead to a reduction in computational time for these mixed models, and may be the only practical approach if there are numerous random coefficients. Third, the random coefficients autoregressive structure is a general panel specification that subsumes all other structures as special cases.

3.1 Experimental Set-up

In the simulation set-up, we consider five choice occasions ($J = 5$) per individual and 500 individuals, with five independent variables per choice occasion. While the number of choice occasions per individual and the number of independent variables per choice occasion can be much larger, we use five choice occasions and five independent variables so that the MSL estimation (which entails a five dimensional integral) is manageable in the context of the many MSL runs we undertake in the paper, while also being reasonably realistic of the kinds of panel setting encountered in practice. The intent is to compare the MSL estimation results and the CML results in terms of the ability to recover parameters as well as computational time. For all the datasets generated in the experimental design, the values of each of the five independent variables are drawn from a standard univariate normal distribution. In the subsequent two sections, we discuss the set-up for each of the random coefficients (RC) and the random coefficients autoregressive (RCA) structures in detail.

3.1.1 The RC Structure

In the RC structure, a coefficient vector β_q (specific to each individual) is assumed and is drawn from a multivariate normal distribution with a mean vector of \mathbf{b} ($= 1.5, 1, 2, 1, 2$). We then consider both independent realizations as well as correlated realizations for the coefficient vector β_q (across exogenous variables for each individual q):

$$\mathbf{\Omega}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0 & 0 \\ 0 & 0 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \text{ and } \mathbf{\Omega}_2 = \begin{bmatrix} 1 & -0.5 & 0.25 & 0.75 & 0 \\ -0.5 & 1 & 0.25 & -0.5 & 0 \\ 0.25 & 0.25 & 1 & 0.33 & 0 \\ 0.75 & -0.5 & 0.33 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (19)$$

For each of the above two positive-definite covariance matrices, the random vector realization of $\boldsymbol{\beta}_q$ is applied to the q th individual's choice occasions, and is linearly combined with the corresponding vector of independent variables ($\boldsymbol{\beta}'_q \mathbf{x}_{qj}$). The result is added to an independent standard normal error term draw (ε_{qj}) as in Equation (1) to obtain a value of y_{qj}^* . This is then translated to “observed” values of y_{qj} based on the following pre-specified threshold values: $\psi^1 = 0$, $\psi^2 = 1.0$, $\psi^3 = 2.0$ (thus we assume four outcome levels for the ordinal variable). The above data generation process is undertaken 50 times with different realizations of the $\boldsymbol{\beta}_q$ vector and the error term ε_{qj} to generate 50 different data sets (for each of the $\mathbf{\Omega}_1$ and $\mathbf{\Omega}_2$ specifications).

The MSL and CML estimation procedures are applied to each data set to estimate data specific values of \mathbf{b} , \mathbf{L}_1 ($\mathbf{\Omega}_1 = \mathbf{L}_1 \mathbf{L}'_1$) or \mathbf{L}_2 ($\mathbf{\Omega}_2 = \mathbf{L}_2 \mathbf{L}'_2$) (as appropriate), and the threshold vector $\boldsymbol{\psi} = (\psi^1, \psi^2, \psi^3)$.¹¹ Note that \mathbf{L}_1 and \mathbf{L}_2 are the lower Cholesky decompositions of the covariance matrices $\mathbf{\Omega}_1$ and $\mathbf{\Omega}_2$, respectively. We estimate the Cholesky parameters to ensure the positive definiteness of the variance-covariance matrices $\mathbf{\Omega}_1$ and $\mathbf{\Omega}_2$. For the MSL estimation, we use draws from the randomized Halton sequence for the random coefficients vector $\boldsymbol{\beta}_q$, because it is the most commonly used QMC sequence in the literature (Halton, 1960; see Bhat, 2003 for a discussion). Within the context of the Halton draws, we experimented with different kinds of scramblings and randomizations of the Halton sequence. This included the following: (a) scrambling the Halton draws for different dimensions using the Bratten-Weller approach to break correlations across dimensions arising from the periodic cycling of the Halton draws, (b) scrambling the Halton draws using a randomization approach to break correlations across dimensions, (c) randomizing the Halton draws along each dimension by adding a uniform

¹¹ Since the vector \mathbf{x}_{qj} does not include a constant in the simulation set-up, the first threshold ψ^1 is estimable.

random number that still preserves the uniformly distributed and equi-distribution properties of the underlying Halton sequence (referred to as the Tuffin-randomization in the literature; see Tuffin, 1996), and (d) randomizing the assignment of Halton dimensions to the random coefficients (so that, for example, the Halton dimension that is based off the prime number two is assigned to say the first random coefficient in one of the simulation runs, while the same Halton dimension is assigned to a different random coefficient in another simulation run). Our experiments suggested that the best performance was obtained using a procedure that combined Bratten-Weller scrambling with the Tuffin randomization as well as the random assignment of Halton dimensions to coefficients. Finally, while a higher number of draws per individual (based on the combination scrambling/randomization discussed above) generally provided improved results, we used 150 draws per individual, which is about what is typically used in most applications of the MSL procedure for ordered-response models. As we will indicate later, we also undertook the MSL estimation with 250 draws per individual, and the results were not substantially different. To assess and quantify simulation variance, the randomized and scrambled Halton-based simulation procedure is applied to each dataset 10 times with different (independent) randomized Halton draw sequences.

3.1.2 The RCA Structure

For the RCA structure, we generate β_q vectors for each individual q based on the mean vector \mathbf{b} ($= 1.5, 1, 2, 1, 1$) and a covariance matrix given by $\mathbf{\Omega}_2$ (we use the more general non-diagonal covariance matrix used in the RC structure for the RCA structure). The rest of the procedure is the same as the RC structure, except for the generation of the standard normal error terms (ε_{qj}). Specifically, these error terms are now serially correlated for each individual q . We assume that this serial correlation gets manifested in the last four of the five observations for each individual, with the first observation error term ε_{q1} for each individual randomly drawn from a standard normal distribution. That is, ε_{qj} ($j = 1, 2 \dots J$) is generated for each individual q as follows:

$$\varepsilon_{qj} = \begin{cases} \eta_{q1} \sim N(0, 1) \text{ for } j = 1 \\ \rho \varepsilon_{qj-1} + \left(\sqrt{1 - \rho^2}\right) \eta_{qj}, \quad \eta_{qj} \sim N(0, 1) \text{ for } j \geq 2. \end{cases} \quad (20)$$

The resulting correlation matrix of ε_{qj} is \mathbf{R}_q which is a $J \times J$ dimension matrix with its jg^{th} element being $\rho^{|j-g|}$, $g, j = 1, 2, \dots, J$.¹² In the current paper, we undertake the simulation exercise for low autocorrelation ($\rho = 0.3$) and high autocorrelation ($\rho = 0.7$). For each of these correlation values, error term realizations of $\boldsymbol{\varepsilon}_q = (\varepsilon_{q1}, \varepsilon_{q2}, \dots, \varepsilon_{qJ})'$ are drawn and used to generate data for the RCA structure. To examine the impact of different magnitudes of the autoregressive correlation parameter, the process is undertaken 20 times with different realizations of the $\boldsymbol{\beta}_q$ and $\boldsymbol{\varepsilon}_q$ vectors to generate 20 different data sets (for each value of ρ of 0.3 and 0.7). We used fewer data sets for the RCA case compared to the RC case because of the substantially increased computational cost for the RCA case with the MSL method. Further, for the MSL approach, it was too expensive to run estimations multiple times with the same data using 150 Halton draws per individual. So, for the two non-diagonal RCA structures (with low and high autocorrelation), we applied the average magnitude of simulation standard error, (as a percentage of the non-simulation adjusted finite sample standard error) as obtained from the non-diagonal RC structure to the non-simulation adjusted finite sample standard errors of the RCA non-diagonal structures to estimate the finite sample standard error for each parameter.¹³ The simulation errors estimated in this way for the RCA non-diagonal cases may be regarded as lower bounds of the true simulation errors, since the simulation errors are expected to be higher for the case with autocorrelation than without autocorrelation

The MSL and CML estimation procedures are applied to each data set to estimate data-specific values of \mathbf{b} , \mathbf{L}_2 , $\boldsymbol{\psi}$, and ρ . To ensure that $0 < \rho < 1$, we re-parameterized ρ as follows:

¹² This is the autoregressive structure of order 1. One can also use more complicated autoregressive structures of order p for the error terms, or use more general structures for the error correlation. For instance, while we focus on a time series context, in spatial contexts related to ordered-response modeling, Bhat *et al.* (2010b) developed a specification where the correlation in physical activity between two individuals may be a function of several measures of spatial proximity and adjacency.

¹³ To be precise, for each data set s for the RC model, we computed the simulation standard deviation for each parameter as the standard deviation in the estimated values across the independent draws. Next, for each parameter, we took the mean of the simulation standard deviations across the different data sets, and computed this simulation standard deviation as a percentage of the non-simulation adjusted standard deviation of the mean parameter value (within each data set s) across the different datasets. This percentage value of the simulation standard error to the non-simulation adjusted finite sample standard error from the RC model for each parameter was then applied to the standard deviation (across the 20 datasets) of the corresponding parameter from each RCA model to obtain the overall finite sample standard error for that RCA model.

$\rho = 1/[1 + \exp(-\Delta)]$. The MSL estimation procedure uses the GHK simulation procedure using 150 draws per individual of the randomized and scrambled Halton sequence (see Bhat *et al.*, 2010a for a discussion of the GHK simulator in the context of ordered response models).¹⁴ In terms of parameter search, the MAXLIK module in the GAUSS matrix programming software is used after coding both the MSL function and the corresponding analytic gradient function. The BFGS secant algorithm is employed for determining the direction of search and the STEPBT line search method is adopted for determining step length. We also experimented with other algorithms and line search methods, but none provided any computational or convergence benefits. Indeed, almost all of the other combinations of algorithm and line search methods took longer than the BFGS-STEPBT combination used in the paper.

4. PERFORMANCE COMPARISON BETWEEN THE MSL AND CML APPROACHES

In this section, we first identify a number of performance measures and discuss how these are computed for the MSL approach and the CML approach. The subsequent sections present the simulation and computational results.

4.1 Performance Measures

The steps discussed below for computing performance measures are for a specific covariance matrix pattern. For the RC model, we consider zero covariance across the random coefficients (diagonal covariance specification) and non-zero covariance across the random coefficients (non-diagonal covariance specification). For the RCA structure, we retain the non-diagonal covariance matrix specification of the RC structure and then consider two correlation patterns, corresponding to the autoregressive correlation parameter values of 0.3 and 0.7.

¹⁴ We use a Halton-based GHK procedure rather than a standard pseudo-random-based GHK procedure, because Halton-based procedures are faster (see Bhat *et al.*, 2010a and Heiss, 2010). Other possibilities to improve upon the standard GHK simulator, all of which are based on more efficiently computing the rectangular multivariate normal cumulative distribution function evaluations within the GHK simulator, include the use of (1) efficient importance sampling techniques as proposed by Liesenfeld and Richard (2010), (2) decomposition techniques as proposed by Huguenin *et al.* (2009), (3) the multivariate sparse grid (SG) Gaussian quadrature approach of Heiss and Winschel (2008), (4) The SG approach combined with an efficient importance sampling (EIS) approach as employed by Heiss (2010), and (5) the Laplace Approximation approach combined with a Gauss-Hermite quadrature technique of Joe (2008). In the current paper, we do not undertake a comprehensive examination of the performance of these alternative GHK simulators (this is beyond the scope of this paper). Rather, we use the Halton-based GHK procedure for the MSL estimation.

MSL Approach

- (1) Estimate the MSL parameters for each data set s ($s = 1, 2, \dots, 50$) and for each of 10 independent draws (the 10 independent draws are different across data sets), and obtain the time to get the convergent values and the standard errors. Note combinations for which convergence is not achieved. Everything below refers to cases when convergence is achieved. Obtain the mean time for convergence (TMSL) and standard deviation of convergence time across the converged runs and across all data sets (the time to convergence includes the time to compute the covariance matrix of parameters and the corresponding parameter standard errors). All estimations are started with the true parameter values as the starting values. While multiple computers had to be used for the many different runs undertaken in this paper, all the run times were carefully scaled to the equivalent time on a desktop computer with 3.60GHz Quadcore processor and 8.00 GB of RAM. The scaling was based on extensive experimentation on different computers.
- (2) For each data set s , compute the mean estimate for each model parameter across the draws. Label this as MED, and then take the mean of the MED values across the data sets to obtain a mean estimate. Compute the finite sample absolute percentage bias (APB) as:
$$APB = \left| \frac{\text{mean estimate} - \text{true value}}{\text{true value}} \right| \times 100$$
¹⁵
- (3) Next, compute the standard deviation (about the mean estimate obtained in the earlier step) of the estimates of the parameter values across the data sets and across draws within each data set. Call this FSSE and label this as the finite sample standard error.
- (4) For each data set s and draw combination, estimate the asymptotic standard errors (s.e.) of parameters (using the sandwich estimator).
- (5) For each data set s , compute the median asymptotic standard error for each model parameter across the draws. Call this MSED, and then take the mean of the MSED values across the data sets and label this as the asymptotic standard error (essentially this is the standard error of the distribution of the estimator as the sample size gets large).
- (6) Compute the absolute percentage bias of the asymptotic standard error for each parameter relative to the corresponding finite sample standard error.

¹⁵ If the true parameter value is zero, the APB value is computed by dividing the mean estimate by the value of 1 in the denominator, and multiplying by 100.

CML Approach

- (1) Estimate the CML parameters for each data set s and obtain the time to get the convergent values (including the time to obtain the Godambe matrix-computed asymptotic covariance matrix and corresponding standard errors). Determine the mean time for convergence (TCML) across the S data sets.¹⁶
- (2) Compute the mean estimate for each model parameter across the R data sets. Compute the absolute percentage bias (APB) as in the MSL case.
- (3) Next, compute the standard deviation of parameters across the data sets about the mean estimate obtained in Step 3. Call this FSSE and label this as the finite sample standard error.
- (4) For each data set s , estimate the asymptotic standard errors (s.e.) (using the Godambe estimator).
- (5) Compute the mean of the asymptotic standard errors for each model parameter across the R data sets and label this as the asymptotic standard error
- (6) Compute the absolute percentage bias of the asymptotic standard error for each parameter relative to the corresponding finite sample standard error.

5. RESULTS

5.1 RC Structure

Tables 1a and 1b provide the results of the RC structure for the diagonal and non-diagonal cases, respectively. The tables provide the true value of the parameters (second column), followed by the maximum simulated likelihood (MSL) estimation results and the composite marginal likelihood (CML) estimation results.

The Diagonal Covariance Matrix Case

The columns under “parameter estimates” in Table 1a provides the mean parameter estimates across data sets and runs, as well as the absolute percentage bias (APB) values. These results indicate that both the MSL and CML methods perform reasonably well in recovering parameters. Specifically, the APB values for the parameters range from 1.06% to 16.84 % for the MSL method and 0.71% to 10.38 % for the CML method. The mean APB using 150 Halton draws per individual in the MSL case is 7.75% (see the row of the table labeled “Overall mean value across

¹⁶ The CML estimator always converged in our simulations, unlike the MSL estimator.

parameters” and the column titled “absolute percentage bias”), while the mean APB from the CML approach is lower at 5.51%. There are two interesting observations from the mean estimate and APB values. First, the MSL estimation underestimates parameter values (the mean estimate values from the MSL are consistently lower than the true values), while the CML estimation overestimates parameter values (except, for the first parameter). This is the case for both the non-scale parameter estimates (the b and ψ estimates) as well as the scale parameters (as captured in the l values). Thus, when normalized for scale, the results indicate that the net effects of explanatory variables are about the same from the MSL and CML estimations. Second, the APB values for the Cholesky parameters (*i.e.*, the l values in the table, which in this diagonal case are the standard deviations of the distributions of each of the five random coefficients) are generally somewhat higher relative to the threshold parameters (*i.e.*, the ψ values in the table) and the mean values of the distributions of the β parameter vector (*i.e.*, the b values in the table), especially so in the MSL estimation. Also, there is more variation in the APB values among the Cholesky parameters than among the b and ψ values in both the MSL and CML estimations. This is perhaps because the Cholesky parameters enter the likelihood function in a more complex non-linear fashion than other parameters, leading to a relatively flat log-likelihood function for different values of standard deviations of the random coefficients and more difficulty in accurately recovering these standard deviation parameters.

The finite sample standard error values of the parameters indicate good efficiency of both the MSL and CML estimators, with the finite sample standard error ranging from 6-27% of the mean values of the MSL estimator and from 11-23% of the mean values of the CML estimator.¹⁷ The magnitudes of the finite sample standard error values are certainly lower in the MSL (mean finite sample standard error of 0.11) compared to the CML (mean finite sample standard error of 0.17). However, note also that this direct comparison of the finite sample standard error values is somewhat deceptive, because of the underestimation (overestimation) in recovering the true values of the parameters in the MSL (CML). This translates to consistently lower values of the finite sample standard error estimates from the MSL approach relative to the CML approach. Another observation from the finite sample standard error estimates is that these estimates (as a

¹⁷ We do not include the first parameter ψ^1 in computing these ranges, because the true value of this parameter is zero, and the mean estimate is also very close to zero. Thus, percentages taken with respect to the mean estimate will be very high.

percentage of the mean estimates) are generally higher for the Cholesky parameters relative to the other parameters, reinforcing the finding earlier that the Cholesky parameters are more difficult to recover than other parameters.

The finite sample standard errors and the asymptotic standard errors obtained using the sandwich estimator in the MSL method are quite close, with a mean APB value of 8.00%. Similarly, the finite sample standard errors and the asymptotic standard errors obtained using the Godambe matrix in the CML method are also close, with a mean APB of 6.09%. These results indicate that the asymptotic formulae are performing accurately in the simple random coefficients case with a diagonal covariance matrix and no autoregressive error. The final column of Table 1a provide the relative finite sample efficiency factors between the MSL and CML estimators. As indicated earlier, the CML approach should lose some efficiency relative to the full maximum likelihood (ML) approach, because the CML approach compounds pairs of observations from the same individual, and does not consider all the panel observations from the same individual simultaneously. However, when we move from the ML approach to the maximum simulated likelihood (MSL) approach, there is again a drop in efficiency because of simulation noise. Thus, one does not know *a priori* whether the MSL estimator will be more efficient than the CML estimator, and, if so, by how much. No theoretical results are derivable, and one has to consider this as an empirical issue. The ratio of the finite sample standard errors provide an empirical estimate of the relative finite sample efficiency of the CML compared to the MSL. Relative efficiency values lower than one indicate a lower CML efficiency relative to the MSL, while values higher than one indicate that the CML is more efficient than the MSL. The results in the final column do indicate that the efficiency of the CML approach is about 51-99% (mean of 65%) of the MSL approach for this simple random coefficients case with a diagonal covariance matrix and no autoregressive error. The time to convergence for the MSL estimation has a mean value of 4.37 minutes with a standard deviation of about 1.36 minutes. On the other hand, the time to convergence for the CML estimation has a mean value of 0.55 minutes with a standard deviation of about 0.1 minutes. This indicates that the CML method is about eight times faster than the MSL estimation. Further, note that the CML method is actually more effective than suggested by this factor of six, because it produces more accurate estimates than the MSL estimates. In fact, it took about 250 Halton draws per individual to reach about the same level of mean APB value as for the CML approach, and the mean time for convergence

with 250 Halton draws is about 6.85 minutes, suggesting a time efficiency factor of over 12 for the CML method relative to the MSL method. Also, while the APB improved as we moved from 150 Halton draws to 250 Halton draws in the MSL approach, there was surprisingly little change in the simulation error. Further, with 250 draws, the asymptotic standard errors increased because there was less underestimation in recovering parameters, with the net result that the relative efficiency of the CML actually marginally improved relative to the MSL method. One other problem we found even in this simple random coefficients MSL estimation was that 36 of the 500 runs did not converge (that is, 7.2% of total runs did not converge). This confirms that, even for the low to moderate dimensions of integration, numerical simulators can lead to numerical instability and convergence problems. On the other hand, no convergence issues whatsoever were encountered with the CML estimation.

Overall, for the specific case of the panel random-coefficients ordered-response model (with no autoregressive error structure), the results here indicate that the MSL estimator is more efficient than the CML estimator, but also that the CML estimator has a computational cost efficiency gain by a factor of about 8. Of course, the CML has the advantage of reproducibility of results, since it is simulation-free (and, as indicated earlier, the simulation errors are not insignificant). It may be expected that, as the number of random coefficients increase, the econometric efficiency gains of the MSL will slip and the computational efficiency gains of the CML will increase. Besides, with large data sets and several specifications to potentially test, even a computational efficiency gain of 8 can be quite substantial. Of course, as we will see in the rest of this paper, as soon as one introduces more realistic and flexible specifications (such as non-diagonal random coefficients and autoregressive error structures), the MSL estimation approach shows little to no econometric efficiency gains over the CML approach and/or literally become infeasible in practice from a computation cost standpoint.

The Non-Diagonal Covariance Matrix Case

The results in Table 1b provide information on the true values, the mean estimates, and the standard errors for the threshold parameters (the ψ parameters), the mean values of the distribution of the β parameter vector (*i.e.*, the b values in the table) and the Cholesky-decomposed parameters characterizing the covariance matrix of the β parameter vector (*i.e.*, the l values in the table).

As in the diagonal case, both the MSL and CML approaches do very well in recovering the parameters, with the APB values ranging from 0.01% to 28.8% (mean of 6.53%) for the MSL and from 0.04% to 29% (mean of 5.83%) for the CML approach. Though these are very good recovery statistics, the MSL method, in general, continues to under-estimate the magnitudes of parameters, while the CML method over-estimates the magnitudes of parameters. Also, similar to the diagonal case, there is more stability in the APB values across the ψ and b parameters than for the Cholesky parameters (the l values), with the APB values for some of the Cholesky elements being rather high. However, these high APB values are also somewhat deceptive, because the estimated values of the Cholesky parameters are not too far away from the true values. But the small magnitudes of the true Cholesky parameter values tend to inflate the APB values. For instance, the highest APB of 29% for the CML method is for the l_{42} parameter, even though the estimated value of -0.1862 is not far from the true value of -0.1443. Also, with a limited sample size and several Cholesky parameters to estimate, one would only expect a little more difficulty in accurately and precisely recovering the Cholesky parameters with finite samples. This is also noticeable in the finite sample standard errors that are in the order of about 10% of the mean estimates for the ψ and b parameters (in both the MSL and CML cases), but much higher for the l parameters. Between the MSL and the CML estimators, the finite sample standard errors are more similar in this non-diagonal case than in the diagonal case of the previous section, with the mean finite sample standard error being 0.17 in the MSL case and 0.19 in the CML case. This is because the MSL provides estimates that are closer to the true values, and to the values from the CML estimation.

For both the MSL and CML estimators, the asymptotic standard errors are somewhat smaller compared to the finite sample standard errors, suggesting that the sample size used in the simulation is somewhat smaller than what is needed in this non-diagonal case for the asymptotic approximations to hold accurately. However, it is clear that the CML estimator has less bias than the MSL estimator.

The final column of the table indicates that the relative efficiency values are close to 1, with some values being higher than 1. The mean relative efficiency value is 0.88, indicating that the CML estimator is not losing much efficiency compared to the MSL estimator.¹⁸ However, it

¹⁸ We also examined the results with 250 Halton draws per individual. However, there was little reduction in the simulation errors. To be specific, the mean simulation error as a percentage of the non-simulation adjusted standard

is also interesting to notice that the CML has a consistently lower efficiency for the non-Cholesky parameters, while the MSL appears to have a lower efficiency for the Cholesky parameters. This may also explain the generally higher APB values for the CML (relative to the MSL) for the non-Cholesky parameters and the generally (but not always) higher APB values for the MSL (relative to the CML) for the Cholesky parameters.¹⁹

The mean computational time for the MSL method is about 10.52 minutes (with a standard deviation of 3.72 minutes) compared to 2.28 minutes (with a standard deviation of 0.48 minutes) for the CML method. So, the CML method is again about 5 times faster compared to the MSL method, for about the same level of accuracy in recovering parameters. At the same time, the econometric efficiency of the CML estimator is as good as the MSL estimator. Besides, 54 of the 500 runs did not converge in the MSL approach, with no such problems with the CML approach.

5.2 RCA Structure

As discussed earlier, the simulations for the case of random coefficients with an autoregressive error structure (*i.e.*, the RCA structure) is undertaken with only 20 datasets. Further, we did not undertake multiple runs for each of the 20 datasets because of the extremely high computational costs of doing so. The simulation errors were estimated based on those obtained from the RC structure. Tables 2a and 2b provide the results for the RCA non-diagonal case with low and high auto-correlations respectively.

Low Autoregressive Correlation Case

The mean APB values are of about the same order in the MSL and the CML methods with the APB ranging from 0.10% to 15.56% (mean of 4.48%) for the MSL and ranging from 0.17% to 23.3% (mean of 6.85%) for the CML. Overall, the parameters are recovered well in both the MSL and CML cases, with a slight edge for the MSL over the CML. The autoregressive

deviation of the mean parameter value (within each data set s) across the different datasets was still about 58% with 250 draws compared to 69% with 150 draws. The mean relative efficiency turned out to be 0.88 with 250 draws, indicating that the relative efficiency between CML and MSL does not change even if the number of draws is increased. Further, the APB values did not improve very substantially -- the mean APB value turned out to be 5.45% with 250 draws compared to 6.53% with 150 draws.

¹⁹ The lower efficiency of parameters leads to a higher scatter of parameters across a finite number of data sets, which can then lead to a higher APB.

correlation parameter ρ is estimated reasonably well both in MSL and CML methods with APB values of 2.33% and 7.21%, respectively.

The estimated finite sample standard error for the MSL approach (mean value of 0.16) and the finite sample standard errors from the CML approach (mean value of 0.19) are quite close to one another. With regard to the accuracy of the asymptotic formulae, the differences between the finite sample standard errors and the asymptotic standard errors in the MSL method are similar to those observed in the non-diagonal RC case. However, in the CML approach, the APB values are quite high for some of the Cholesky parameters, especially relative to the non-diagonal case.

The relative efficiency values (last column of table) range from 57% to 127%, with a mean value of 87%. This indicates relatively little overall econometric efficiency loss in using the CML approach relative to the MSL. Note also that because we had to estimate the simulation errors based on the results from the random coefficients (RC) case, the relative efficiency values in Table 2 may be considered as lower bounds, implying that the mean relative efficiency value is likely to be 87% or higher. At the same time, the mean time to convergence is 228.6 minutes or about 4 hours (standard deviation of 42.0 minutes) for the MSL compared to only about 2.19 minutes for the CML method (with a standard deviation of 0.54 minutes). This is a phenomenal computation efficiency leap, with the CML method being about 104 times faster than the MSL method. We observed 100% convergence rates for both the MSL and CML methods.

High Autoregressive Correlation Case

The APB for the MSL approach ranges from 0.44% to 26.2% (mean of 4.85%), while the CML APB ranges from 0.59% to 17.95% (mean of 6.75%). While both of these APBs are of about the same order, the MSL approach does provide a marginally better APB in the high correlation case as well. However, in general, the ability to recover parameters does not seem to be affected at all by whether there is low correlation or high correlation. With specific regard to the autoregressive correlation parameter (ρ), the result indicates that the approaches the CML approach, in particular, recover this parameter very well in the high correlation case relative to the low correlation case (an associated APB of 0.93% in the high correlation case, compared to a corresponding APB of 7.21% in the low correlation case). This is perhaps because the correlation parameter needs to be particularly strong before it starts having any substantial effects on the

log-likelihood function value. Essentially, the log-likelihood function can be relatively flat at low correlation, leading to more difficulty in accurately recovering the low correlation parameter. But, at a high correlation level, the log-likelihood function shifts considerably in value with small shifts in the correlation value, allowing it to be recovered accurately. Why this does not play out in the MSL case is an open question. It is possible that, at high correlations, there is considerable instability in the search direction and the convergence process in the MSL approach. This is reinforced by the fact that, while all the 20 runs converged (100% convergence rate) in the MSL approach in the low correlation case, only 14 of the 20 runs converged (70% convergence rate) in the MSL approach in the high correlation case.²⁰

The APB values for the asymptotic standard error once again are in the reasonable range, though it is now the ML estimator that seems to have some large APB values for some of the Cholesky parameters. The CML asymptotic standard error APB values are much smaller than the low autoregressive correlation case. The relative efficiency values in the last column reveal that there is an even lower efficiency loss with the CML approach relative to the MSL approach in this high correlation case compared to the low correlation case. The MSL runs that converged have a mean time to convergence of 200.05 minutes (more than 3.5 hours) relative to only 2.04 minutes for the CML method; that is, the CML method is about 98 times faster than the MSL method. Note also that there is a huge standard deviation in the time to convergence of the MSL method, which is consistent with the convergence-related instability problems discussed earlier. In cases with more than five random coefficients and more general auto-regressive structures than the simple first order structure considered in this study, the convergence problems and the high computational times of the MSL make it less appealing. The CML method, on the other hand, should be able to accommodate such panel structures with relative ease.

6. SUMMARY AND CONCLUSIONS

This paper focuses on panel ordered-response model structures, and compares the performance of the maximum simulated likelihood (MSL) estimation approach with that of the composite marginal likelihood (CML) estimation approach. The panel structures considered in the paper

²⁰ We undertook an independent simulation exercise to see if the same convergence problems held up with a diagonal instead of a non-diagonal covariance structure for the random coefficients in this RCA case. The results showed the same trend as in the non-diagonal case discussed here, with the convergence rates being 100% for the low correlation case compared to only 70% for the high correlation case. Thus, the convergence problem seems quite definitely associated with the correlation level.

include the pure random coefficients (RC) model with no autoregressive error component, as well as the more general case of random coefficients combined with an autoregressive error component. The ability of the MSL and CML approaches to recover the true parameters is examined using simulated datasets.

Overall, the results suggest that the CML method is able to recover the true parameters in all the cases considered in the paper, irrespective of the type of covariance matrix (diagonal versus non-diagonal) of the random coefficients and the level of the autoregressive correlation (low versus high). In fact, the performance of the MSL approach with 150 randomized and scrambled Halton draws and the simulation-free CML approach are about the same order in all cases in terms of the absolute percentage bias (APB) of the parameters estimated. The MSL approach has an edge in terms of econometric efficiency in the context of the pure RC model with a diagonal covariance matrix for the random coefficients, but also suffers from convergence problems even in this simple panel case. The CML approach does not exhibit any convergence problems, and is about eight times faster. Any econometric efficiency gains of the MSL approach vanishes as soon as a non-diagonal covariance matrix of the random coefficients is introduced, due to the presence of large simulation noise in the MSL runs. This continues to hold when an autoregressive error structure is added to the non-diagonal random coefficients structure. Increasing the number of randomized/scrambled Halton draws from 150 per individual to 250 per individual has little impact on these results. Further, when an autoregressive error structure is added, the convergence times of the MSL runs start getting very high, while there is little to no effect on the convergence times for the CML runs. In fact, the CML method is about 100 times faster than the MSL method when there is an autoregressive error component. The MSL approach breaks down in particular when the autoregressive error correlation magnitude is high, with a substantial number of runs failing to converge. At the same time, there is literally no loss in econometric efficiency of the CML estimator compared to the MSL estimator.

While the asymptotic formulae seem to be performing quite accurately for the sample sizes considered in this study in the RC model with diagonal covariance matrix case, there are somewhat larger differences between the asymptotic standard errors and the finite sample standard errors in all the non-diagonal cases (both RC and RCA). This is to be expected because of the larger number of parameters and more complicated structures of the non-diagonal models.

Overall, the asymptotic standard errors appear to be more accurate when the CML estimator is used relative to the MSL estimator.

Future research efforts should consider varying numbers of random coefficients, autoregressive error structures of order higher than one, varying numbers of observations per individual, and non-normal mixing distributions. However, the results in this paper paint a very encouraging picture for the use of the CML pairwise likelihood method for the quick, accurate, and practical estimation of panel ordered-response models with flexible and rich stochastic specifications.

ACKNOWLEDGEMENTS

The author acknowledges the helpful comments of three anonymous reviewers on an earlier version of the paper. This research was partially funded by a Southwest Region University Transportation Center grant. The authors are grateful to Lisa Macias for her help in typesetting and formatting this document.

REFERENCES

- Bartels, R., D. G. Fiebig, and A. van Soest (2006) Consumers and experts: an econometric analysis of the demand for water heaters. *Empirical Economics*, **31** (2), 369-391.
- Beron, K. J. and W. P. M. Vijverberg (2004) Probit in a spatial context: A Monte Carlo analysis. In *Advances in Spatial Econometrics: Methodology, Tools and Applications*, Anselin, L., Florax, R.J.G.M., Rey, S.J. (eds.), Springer-Verlag, Berlin, 169-196.
- Bhat, C. R. (2001) Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model. *Transportation Research Part B*, **35** (7), 677-693.
- Bhat, C. R. (2003) Simulation estimation of mixed discrete choice models using randomized and scrambled Halton sequences. *Transportation Research Part B*, **37** (9), 837-855.
- Bhat, C. R. (2011) The maximum approximate composite marginal likelihood (MACML) estimation of multinomial probit-based unordered response choice models. *Transportation Research Part B* **45** (7), 923-939.
- Bhat, C. R. and R. Sardesai (2006) The impact of stop-making and travel time reliability on commute mode choice. *Transportation Research Part B*, **40** (9), 709-730.
- Bhat, C. R. and R. Sidharthan (2012) A new approach to specify and estimate non-normally mixed multinomial probit models. *Transportation Research Part B*, **46** (7), 817-833.
- Bhat, C. R. and H. Zhao (2002) The spatial analysis of activity stop generation. *Transportation Research Part B*, **36** (6), 557-575.
- Bhat, C. R., C. Varin, and N. Ferdous (2010a) A comparison of the maximum simulated likelihood and composite marginal likelihood estimation approaches in the context of the multivariate ordered response model. In *Advances in Econometrics: Maximum Simulated Likelihood Methods and Applications*, Vol. 26, Greene, W.H., Hill, R.C. (eds.), Emerald Group Publishing Limited, 65-106.
- Bhat, C. R., I. N. Sener, and N. Eluru (2010b) A flexible spatially dependent discrete choice model: formulation and application to teenagers' weekday recreational activity participation. *Transportation Research Part B*, **44** (8-9), 903-921.
- Butler, J. S. and R. Moffitt (1982) A computationally efficient quadrature procedure for the one-factor multinomial probit model. *Econometrica*, **50** (3), 761-764.
- Cox, D. R. and N. Reid (2004) A note on pseudolikelihood constructed from marginal densities. *Biometrika*, **91** (3), 729-737.
- Eluru, N., C. R. Bhat, and D. A. Hensher (2008) A mixed generalized ordered response model for examining pedestrian and bicyclist injury severity level in traffic crashes. *Accident Analysis and Prevention*, **40** (3), 1033-1054.
- Genz, A. and F. Bretz (1999) Numerical computation of multivariate t-probabilities with application to power calculation of multiple contrasts. *Journal of Statistical Computation and Simulation*, **63** (4), 361-378.
- Genz, A. and F. Bretz (2002) Comparison of methods for the computation of multivariate t probabilities. *Journal of Computational and Graphical Statistics*, **11** (4), 950-971.
- Geweke, J. (1991) Efficient simulation from the multivariate normal and student-t distributions subject to linear constraints. *Computer Science and Statistics: Proceedings of the Twenty Third Symposium on the Interface*, 571-578, Foundation of North America Inc., Fairfax.
- Geweke, J. F., M. P. Keane, and D. E. Runkle (1997) Statistical inference in the multinomial multiperiod probit model. *Journal of Econometrics*, **80** (1), 125-165.

- Girard, P. and E. Parent (2001) Bayesian analysis of autocorrelated ordered categorical data for industrial quality monitoring. *Technometrics*, **43** (2), 180-191.
- Godambe, V. P. (1960) An optimum property of regular maximum likelihood estimation. *The Annals of Mathematical Statistics*, **31** (4), 1208-1211.
- Greene, W. H. (2000) *Econometric Analysis* (Fourth ed.), Prentice-Hall Inc., New Jersey.
- Greene, W. H. (2005) Censored data and truncated distributions. In *Palgrave Handbook of Econometrics*, Vol. 1: Econometric Theory, Mills, T.C., Patterson, K. (eds.), Palgrave Macmillan, Hampshire.
- Greene, W. H. and D. A. Hensher (2010) Ordered choices and heterogeneity in attribute processing. *Journal of Transport Economics and Policy*, **44** (3), 331-364.
- Hajivassiliou, V. and D. McFadden (1998) The method of simulated scores for the estimation of LDV models. *Econometrica*, **66** (4), 863-896.
- Halton, J. H. (1960) On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. *Numerische Mathematik*, **2**, 84-90.
- Heiss, F. (2010) The panel probit model: adaptive integration on sparse grids. In *Advances in Econometrics: Maximum Simulated Likelihood Methods and Applications*, Vol. 26, Greene, W.H., Hill, R.C. (eds.), Emerald Group Publishing Limited, 41-64.
- Heiss, F. and V. Winschel (2008) Likelihood approximation by numerical integration on sparse grids. *Journal of Econometrics*, **144** (1), 62-80.
- Hjort, N. L., and C. Varin (2008) ML, PL, QL in Markov chain models. *Scandinavian Journal of Statistics*, **35** (1), 64-82.
- Huguenin, J., F. Pelgrin, and A. Holly (2009) Estimation of multivariate probit models by exact maximum likelihood. Working Paper 0902, University of Lausanne, Institute of Health Economics and Management (IEMS), Lausanne, Switzerland.
- Joe, H. (2008) Accuracy of Laplace approximation for discrete response mixed models. *Computational Statistics and Data Analysis*, **52** (12), 5066-5074.
- Joe, H. and Y. Lee (2009) On weighting of bivariate margins in pairwise likelihood. *Journal of Multivariate Analysis*, **100** (4), 670-685.
- Keane, M. P. (1994) A computationally practical simulation estimator for panel data. *Econometrica*, **62** (1), 95-116.
- Keane, M. P. (1997) Modeling heterogeneity and state dependence in consumer choice behavior. *Journal of Business & Economic Statistics*, **15** (3), 310-327.
- Kuk, A. Y. C. and D. J. Nott (2000) A pairwise likelihood approach to analyzing correlated binary data. *Statistics and Probability Letters*, **47** (4), 329-335.
- Le Cessie, S. and J. C. Van Houwelingen (1994) Logistic regression for correlated binary data. *Applied Statistics*, **43** (1), 95-108.
- Lee, L. (1995) Asymptotic bias in simulated maximum likelihood estimation of discrete choice models. *Econometric Theory*, **11** (3), 437-483.
- Lele, S. R. (2006) Sampling variability and estimates of density dependence: a composite-likelihood approach. *Ecology*, **87** (1), 189-202.
- LeSage, J. P. (2000) Bayesian estimation of limited dependent variable spatial autoregressive models. *Geographical Analysis*, **32** (1), 19-35.
- Lindsay, B. G. (1988) Composite likelihood methods. *Contemporary Mathematics*, **80**, 221-239.
- Liesenfeld, R. and J. F. Richard (2010) Efficient estimation of probit models with correlated errors. *Journal of Econometrics*, **156** (2), 367-376.

- Mardia, K. V., J. T. Kent, G. Hughes, and C. C. Taylor (2009) Maximum likelihood estimation using composite likelihoods for closed exponential families. *Biometrika*, **96** (4), 975-982.
- McFadden, D. and K. Train (2000) Mixed MNL models for discrete response. *Journal of Applied Econometrics*, **15** (5), 447-470.
- Mi, X., T. Miwa, and T. Hothorn (2009) mvtnorm: new numerical algorithm for multivariate normal probabilities. *The R Journal*, **1** (1), 37-39.
- Müller, G. and C. Czado (2005) An autoregressive ordered probit model with application to high frequency financial data. *Journal of Computational and Graphical Statistics*, **14** (2), 320-338.
- Renard, D., G. Molenberghs, and H. Geys (2004) A pairwise likelihood approach to estimation in multilevel probit models. *Computational Statistics & Data Analysis*, **44** (4), 649-667.
- Small, K. A., C. Winston, and J. Yan (2005) Uncovering the distribution of motorists' preferences for travel time and reliability. *Econometrica*, **73** (4), 1367-1382.
- Tuffin, B. (1996) On the use of low discrepancy sequences in Monte Carlo methods. *Monte Carlo Methods and Applications*, **2** (4), 295-320.
- Varin, C. (2008) On composite marginal likelihoods. *AStA Advances in Statistical Analysis*, **92** (1), 1-28.
- Varin, C. and C. Czado (2010) A mixed autoregressive probit model for ordinal longitudinal data. *Biostatistics*, **11** (1), 127-138.
- Varin, C., N. Reid, and D. Firth (2011) An overview of composite marginal likelihoods. *Statistica Sinica*, **21** (1), 5-42.
- Zhao, Y. and H. Joe (2005) Composite likelihood estimation in multivariate data analysis. *The Canadian Journal of Statistics*, **33** (3), 335-356.

LIST OF TABLES

Table 1a Evaluation of Ability to Recover “True” Parameters by the MSL and CML Approaches
–Diagonal Case without Auto-Correlation Random Coefficients (RC) Structure

Table 1b Evaluation of Ability to Recover “True” Parameters by the MSL and CML Approaches
–Non-Diagonal Case without Auto-Correlation Random Coefficients (RC) Structure

Table 2a Evaluation of Ability to Recover “True” Parameters by the MSL and CML Approaches
–Non-Diagonal Case with Low Auto-Correlation Random Coefficients Autoregressive (RCA) Structure

Table 2b Evaluation of Ability to Recover “True” Parameters by the MSL and CML Approaches
–Non-Diagonal Case with High Auto-Correlation Random Coefficients Autoregressive (RCA) Structure

**Table 1a Evaluation of Ability to Recover “True” Parameters by the MSL and CML Approaches
–Diagonal Case without Auto-Correlation Random Coefficients (RC) Structure**

Parameter	True Value	MSL					CML Approach					Relative efficiency $\frac{FSSE_{MSL}}{FSSE_{CML}}$
		Parameter Estimates		Standard error estimates			Parameter Estimates		Standard error estimates			
		Mean Estimate	Absolute percentage bias (APB)	Finite sample standard error ($FSSE_{MSL}$)	Asymptotic standard error ($MASE_{MSL}$)	APB for Asymptotic standard error	Mean Estimate	Absolute percentage bias (APB)	Finite sample standard error ($FSSE_{CML}$)	Asymptotic standard error ($MASE_{CML}$)	APB for Asymptotic standard error	
ψ_1	0.0000	-0.0106	1.06	0.0519	0.0563	8.41	-0.0071	0.71	0.0605	0.0651	7.56	0.86
ψ_2	1.0000	0.9289	7.11	0.0835	0.0810	3.07	1.0522	5.22	0.1412	0.1217	13.82	0.59
ψ_3	2.0000	1.8686	6.57	0.1164	0.1270	9.11	2.1123	5.62	0.2297	0.2153	6.25	0.51
b_1	1.5000	1.3971	6.86	0.1253	0.1140	9.06	1.5949	6.33	0.1889	0.1811	4.14	0.66
b_2	1.0000	0.9327	6.73	0.0891	0.0912	2.39	1.0683	6.83	0.1413	0.1345	4.83	0.63
b_3	2.0000	1.8455	7.72	0.1248	0.1363	9.16	2.1096	5.48	0.2337	0.2287	2.13	0.53
b_4	1.0000	0.9345	6.55	0.0806	0.0855	6.12	1.0671	6.71	0.1423	0.1294	9.08	0.57
b_5	2.0000	1.8388	8.06	0.1240	0.1386	11.77	2.0889	4.45	0.2306	0.2291	0.65	0.54
l_{11}	1.0000	0.9209	7.91	0.1197	0.1131	5.44	1.0705	7.05	0.1831	0.1641	10.38	0.65
l_{22}	1.0000	0.9154	8.46	0.1081	0.1127	4.23	1.0587	5.87	0.1594	0.1649	3.45	0.68
l_{33}	0.7071	0.5880	16.84	0.1618	0.1268	21.66	0.7151	1.13	0.1637	0.1616	1.28	0.99
l_{44}	0.7071	0.6418	9.24	0.0998	0.1070	7.23	0.7799	10.30	0.1594	0.1469	7.84	0.63
l_{55}	1.0000	0.9240	7.60	0.1130	0.1201	6.32	1.0596	5.96	0.1838	0.1694	7.82	0.61
Overall mean value across parameters			7.75	0.11	0.11	8.00		5.51	0.17	0.16	6.09	0.65
Mean Time		4.37					0.55					
Std. Dev. Of Time		1.36					0.10					
% of Runs Converged		92.80%					100%					

**Table 1b Evaluation of Ability to Recover “True” Parameters by the MSL and CML Approaches
–Non-Diagonal Case without Auto-Correlation Random Coefficients (RC) Structure**

Parameter	True Value	MSL					CML Approach					Relative efficiency $\frac{FSSE_{MSL}}{FSSE_{CML}}$
		Parameter Estimates		Standard error estimates			Parameter Estimates		Standard error estimates			
		Mean Estimate	Absolute percentage bias (APB)	Finite sample standard error ($FSSE_{MSL}$)	Asymptotic standard error ($MASE_{MSL}$)	APB for Asymptotic standard error	Mean Estimate	Absolute percentage bias (APB)	Finite sample standard error ($FSSE_{CML}$)	Asymptotic standard error ($MASE_{CML}$)	APB for Asymptotic standard error	
ψ_1	0.0000	-0.0001	0.01	0.0615	0.0575	6.55	-0.0044	0.44	0.0593	0.0649	9.40	1.04
ψ_2	1.0000	0.9510	4.90	0.0829	0.0878	5.90	1.0582	5.82	0.1368	0.1220	10.84	0.61
ψ_3	2.0000	1.9170	4.15	0.1441	0.1457	1.12	2.1142	5.71	0.2495	0.2157	13.58	0.58
b_1	1.5000	1.4308	4.61	0.1346	0.1342	0.31	1.5948	6.32	0.2027	0.1859	8.27	0.66
b_2	1.0000	0.9568	4.32	0.1081	0.1074	0.64	1.0817	8.17	0.1561	0.1405	10.00	0.69
b_3	2.0000	1.8797	6.02	0.1491	0.1672	12.12	2.1153	5.77	0.2624	0.2371	9.62	0.57
b_4	1.0000	0.9437	5.63	0.1026	0.1072	4.49	1.0563	5.63	0.1497	0.1395	6.81	0.68
b_5	2.0000	1.9017	4.92	0.1667	0.1699	1.92	2.0897	4.49	0.2559	0.2346	8.36	0.65
l_{11}	1.0000	0.9493	5.07	0.1337	0.1309	2.10	1.0867	8.67	0.1946	0.1697	12.77	0.69
l_{21}	-0.5000	-0.4696	6.07	0.1352	0.1287	4.80	-0.4987	0.27	0.1360	0.1489	9.47	0.99
l_{22}	0.8660	0.8148	5.91	0.1510	0.1478	2.11	0.9265	6.98	0.1939	0.1814	6.46	0.78
l_{31}	0.2500	0.2474	1.06	0.1531	0.1362	11.05	0.2219	11.24	0.1609	0.1506	6.39	0.95
l_{32}	0.4330	0.4494	3.79	0.1823	0.1517	16.77	0.4497	3.85	0.1512	0.1791	18.43	1.21
l_{33}	0.8660	0.7511	13.27	0.1874	0.1562	16.68	0.8812	1.75	0.1877	0.1851	1.40	1.00
l_{41}	0.7500	0.7121	5.05	0.1328	0.1206	9.19	0.7503	0.04	0.1489	0.1441	3.23	0.89
l_{42}	-0.1443	-0.1128	21.85	0.1890	0.1307	30.87	-0.1862	28.98	0.1729	0.1723	0.32	1.09
l_{43}	0.2367	0.2298	2.93	0.1803	0.1413	21.62	0.2692	13.73	0.1970	0.1737	11.82	0.92
l_{44}	0.6005	0.4277	28.78	0.2621	0.1550	40.86	0.5851	2.56	0.1951	0.2062	5.73	1.34
l_{51}	0.0000	0.0111	1.11	0.1708	0.1302	23.76	-0.0013	0.19	0.1271	0.1496	17.72	1.34
l_{52}	0.0000	-0.0148	1.48	0.1936	0.1519	21.53	-0.0129	1.29	0.2121	0.1879	11.40	0.91
l_{53}	0.0000	-0.0013	0.13	0.2000	0.1590	20.53	-0.0159	1.59	0.2214	0.2034	8.15	0.90
l_{54}	0.0000	0.0151	1.51	0.3550	0.1774	50.02	0.0153	1.53	0.3564	0.3042	14.66	1.00
l_{55}	1.0000	0.8234	17.66	0.2441	0.1626	33.41	0.9088	9.12	0.2983	0.2574	13.70	0.82
Overall mean value across parameters			6.53	0.17	0.14	14.71		5.83	0.19	0.18	9.50	0.88
Mean Time		10.52					2.28					
Std. Dev. Of Time		3.72					0.48					
% of Runs Converged		89.00%					100%					

**Table 2a Evaluation of Ability to Recover “True” Parameters by the MSL and CML Approaches
–Non-Diagonal Case with Low Auto-Correlation Random Coefficients (RCA) Structure**

Parameter	True Value	MSL					CML Approach					Relative efficiency $\frac{FSSE_{MSL}}{FSSE_{CML}}$
		Parameter Estimates		Standard error estimates			Parameter Estimates		Standard error estimates			
		Mean Estimate	Absolute percentage bias (APB)	Finite sample standard error ($FSSE_{MSL}$)	Asymptotic standard error ($MASE_{MSL}$)	APB for Asymptotic standard error	Mean Estimate	Absolute percentage bias (APB)	Finite sample standard error ($FSSE_{CML}$)	Asymptotic standard error ($MASE_{CML}$)	APB for Asymptotic standard error	
ψ_1	0.0000	-0.0010	0.10	0.0590	0.0631	6.91	-0.0017	0.17	0.0618	0.0669	8.30	0.96
ψ_2	1.0000	1.0380	3.80	0.0765	0.0973	27.22	1.0639	6.39	0.1347	0.1290	4.23	0.57
ψ_3	2.0000	2.0614	3.07	0.1638	0.1587	3.10	2.1092	5.46	0.2224	0.2282	2.61	0.74
b_1	1.5000	1.5298	1.99	0.1242	0.1406	13.20	1.5649	4.33	0.1915	0.1908	0.37	0.65
b_2	1.0000	1.0602	6.02	0.1034	0.1129	9.23	1.0813	8.13	0.1437	0.1487	3.51	0.72
b_3	2.0000	2.0514	2.57	0.1598	0.1781	11.48	2.1047	5.23	0.2478	0.2466	0.47	0.64
b_4	1.0000	1.0373	3.73	0.1158	0.1113	3.88	1.0555	5.55	0.1391	0.1470	5.70	0.83
b_5	2.0000	2.0499	2.50	0.2013	0.1789	11.12	2.0794	3.97	0.2344	0.2482	5.88	0.86
l_{11}	1.0000	1.0495	4.95	0.1335	0.1294	3.05	1.0945	9.45	0.1613	0.1734	7.50	0.83
l_{21}	-0.5000	-0.5171	3.42	0.1330	0.1294	2.73	-0.5555	11.10	0.1358	0.1550	14.10	0.98
l_{22}	0.8660	0.9199	6.23	0.1747	0.1414	19.04	0.9205	6.29	0.1788	0.2097	17.30	0.98
l_{31}	0.2500	0.2250	10.01	0.1602	0.1273	20.55	0.2186	12.54	0.1559	0.1508	3.30	1.03
l_{32}	0.4330	0.4990	15.23	0.1375	0.1467	6.68	0.5340	23.33	0.1919	0.2263	17.91	0.72
l_{33}	0.8660	0.8804	1.66	0.1500	0.1444	3.68	0.8912	2.91	0.1784	0.2517	41.07	0.84
l_{41}	0.7500	0.7474	0.35	0.1478	0.1169	20.96	0.7421	1.05	0.1584	0.1444	8.85	0.93
l_{42}	-0.1443	-0.1218	15.56	0.2020	0.1403	30.53	-0.1336	7.45	0.1586	0.1713	8.03	1.27
l_{43}	0.2367	0.2411	1.85	0.1742	0.1374	21.10	0.2536	7.13	0.2009	0.1797	10.57	0.87
l_{44}	0.6005	0.5480	8.75	0.1979	0.1519	23.21	0.5371	10.56	0.1971	0.1950	1.06	1.00
l_{51}	0.0000	0.0163	1.63	0.1203	0.1305	8.48	0.0139	1.39	0.1171	0.1505	28.53	1.03
l_{52}	0.0000	0.0295	2.95	0.1896	0.1520	19.79	0.0230	2.30	0.2209	0.2321	5.08	0.86
l_{53}	0.0000	-0.0208	2.08	0.2337	0.1644	29.64	0.0022	0.22	0.2297	0.3648	58.83	1.02
l_{54}	0.0000	0.0578	5.78	0.4168	0.2501	40.01	0.0408	4.08	0.3646	0.3520	3.45	1.14
l_{55}	1.0000	0.9901	0.99	0.2202	0.1708	22.45	0.8187	18.13	0.3455	0.5906	70.96	0.64
ρ	0.3000	0.2930	2.33	0.1269	0.1097	13.59	0.2784	7.21	0.1574	0.1308	16.88	0.81
Overall mean value across parameters			4.48	0.16	0.14	15.49		6.85	0.19	0.21	14.35	0.87
Mean Time		228.59					2.19					
Std. Dev. Of Time		42.00					0.54					
% of Runs Converged		100.0%					100%					

**Table 2b Evaluation of Ability to Recover “True” Parameters by the MSL and CML Approaches
–Non-Diagonal Case with High Auto-Correlation Random Coefficients (RCA) Structure**

Parameter	True Value	MSL					CML Approach					Relative efficiency $\frac{FSSE_{MSL}}{FSSE_{CML}}$
		Parameter Estimates		Standard error estimates			Parameter Estimates		Standard error estimates			
		Mean Estimate	Absolute percentage bias (APB)	Finite sample standard error ($FSSE_{MSL}$)	Asymptotic standard error ($MASE_{MSL}$)	APB for Asymptotic standard error	Mean Estimate	Absolute percentage bias (APB)	Finite sample standard error ($FSSE_{CML}$)	Asymptotic standard error ($MASE_{CML}$)	APB for Asymptotic standard error	
ψ_1	0.0000	0.0217	2.17	0.0791	0.0674	14.83	0.0130	1.30	0.0776	0.0722	6.98	1.02
ψ_2	1.0000	1.0429	4.29	0.1056	0.0944	10.59	1.0704	7.04	0.1417	0.1212	14.46	0.75
ψ_3	2.0000	2.0359	1.79	0.1508	0.1478	2.03	2.1011	5.06	0.2255	0.2054	8.91	0.67
b_1	1.5000	1.5560	3.73	0.1151	0.1353	17.49	1.5819	5.46	0.1774	0.1764	0.56	0.65
b_2	1.0000	1.0298	2.98	0.1088	0.1056	2.96	1.0724	7.24	0.1415	0.1334	5.70	0.77
b_3	2.0000	2.0267	1.34	0.1690	0.1650	2.34	2.0963	4.82	0.2416	0.2243	7.17	0.70
b_4	1.0000	1.0487	4.87	0.0905	0.1070	18.21	1.0644	6.44	0.1092	0.1356	24.22	0.83
b_5	2.0000	2.0477	2.38	0.1456	0.1683	15.62	2.0914	4.57	0.2220	0.2263	1.96	0.66
l_{11}	1.0000	1.0681	6.81	0.1181	0.1221	3.41	1.1077	10.77	0.1636	0.1611	1.55	0.72
l_{21}	-0.5000	-0.4956	0.87	0.1262	0.1224	3.05	-0.5528	10.55	0.1361	0.1484	9.05	0.93
l_{22}	0.8660	0.8905	2.83	0.1412	0.1266	10.39	0.9058	4.60	0.1842	0.1713	7.03	0.77
l_{31}	0.2500	0.2489	0.44	0.1581	0.1206	23.68	0.2239	10.44	0.1509	0.1479	2.00	1.05
l_{32}	0.4330	0.4906	13.30	0.2046	0.1367	33.19	0.5076	17.24	0.1827	0.1809	1.00	1.12
l_{33}	0.8660	0.8153	5.85	0.2405	0.1382	42.55	0.8812	1.76	0.1772	0.1810	2.17	1.36
l_{41}	0.7500	0.7606	1.41	0.1021	0.1121	9.85	0.7588	1.18	0.1238	0.1375	11.03	0.82
l_{42}	-0.1443	-0.1065	26.20	0.1796	0.1312	26.93	-0.1203	16.64	0.1916	0.1699	11.34	0.94
l_{43}	0.2367	0.2387	0.86	0.1777	0.1511	14.96	0.2792	17.95	0.2051	0.1696	17.33	0.87
l_{44}	0.6005	0.5586	6.97	0.1648	0.1366	17.11	0.5504	8.35	0.1875	0.1900	1.32	0.88
l_{51}	0.0000	0.0787	7.87	0.1576	0.1261	19.99	0.0546	5.46	0.1255	0.1478	17.76	1.26
l_{52}	0.0000	0.0304	3.04	0.2347	0.1451	38.17	0.0059	0.59	0.2289	0.1859	18.78	1.03
l_{53}	0.0000	-0.0837	8.37	0.2427	0.1917	21.01	-0.0467	4.67	0.2072	0.2092	0.95	1.17
l_{54}	0.0000	0.0291	2.91	0.4836	0.2389	50.60	0.0209	2.09	0.3413	0.3239	5.10	1.42
l_{55}	1.0000	0.9710	2.90	0.1899	0.1676	11.72	0.9299	7.01	0.2772	0.2812	1.44	0.68
ρ	0.7000	0.6842	2.25	0.0808	0.0735	8.95	0.7065	0.93	0.0853	0.0959	12.37	0.95
Overall mean value across parameters			4.85	0.17	0.13	17.49		6.76	0.18	0.17	7.92	0.92
Mean Time		200.05					2.04					
Std. Dev. Of Time		46.08					0.70					
% of Runs Converged		70%					100%					